Solution Guide

Dell and Emulex: A lossless 10Gb Ethernet iSCSI SAN for VMware vSphere 5

Solution Implementer's Series™

iSCSI over Data Center Bridging (DCB) solution

EMULEX

Table of contents

Executive summary....3

iSCSI over DCB....3
- How best to integrate an iSCSI SAN?....4
- What does iSCSI over DCB provide?....5
- iSCSI configurations today....6

Configuring iSCSI over DCB....9
- Configuring the network switch....11
- Configuring the storage array....12
- Emulex OneConnect® OCe11100 adapters....13

Validation and troubleshooting....17
- Validating the iSCSI adapter....17
- Obtaining link status....19
- Testing the network switch....20

Conclusion....20

Appendix A- Bill of materials....21

Appendix B – Network switch command line configuration....22

Appendix C- Emulex OneCommand OCe11102-IM adapter configuration....27

For more information....28

Executive summary

As part of the VMware Partner Verified and Supported Products (PVSP) program, Emulex and Dell have tested and validated an iSCSI over Data Center Bridging (DCB) solution for VMware vSphere. This technical document outlines the solution developed by Emulex and Dell, which included a Dell EqualLogic PS6010 iSCSI SAN array with a Dell PowerEdge R710 server using Emulex OCe11102-IM adapters in a 10Gb Ethernet network. The solution featured a converged infrastructure in which a network switch configured for Data Center Bridging (DCB) was able to support both network traffic and lossless iSCSI traffic.

Converged networks are becoming more acceptable in the datacenter. This technology is often used by enterprise customers with greenfield datacenters that need to maintain application performance and also focus on service level agreements; alternatively, converged networks may be attractive in growing datacenters where new technologies are required just to stay current and within budget.

This document provides an in-depth look at iSCSI over DCB and explains how to configure and set up the environment based on best practices for a converged infrastructure.

Intended audience: This document is intended for system engineers, VMware administrators and VMware administrators, SAN administrators and networking engineers.

iSCSI over DCB

So what is iSCSI over DCB? DCB was created to provide enhancements for LAN traffic – in part, to eliminate data losses due to overflowing queues and to provide the capability to allocate specific bandwidths on links. The result has been the introduction of a set of new networking standards, which include the following:

- **Priority Flow Control (PFC) or IEEE 802.1Qbb** – Provides link-level flow control that can be managed independently for each frame priority (as shown in Figure 1), ensuring there are no losses when a DCB network becomes congested

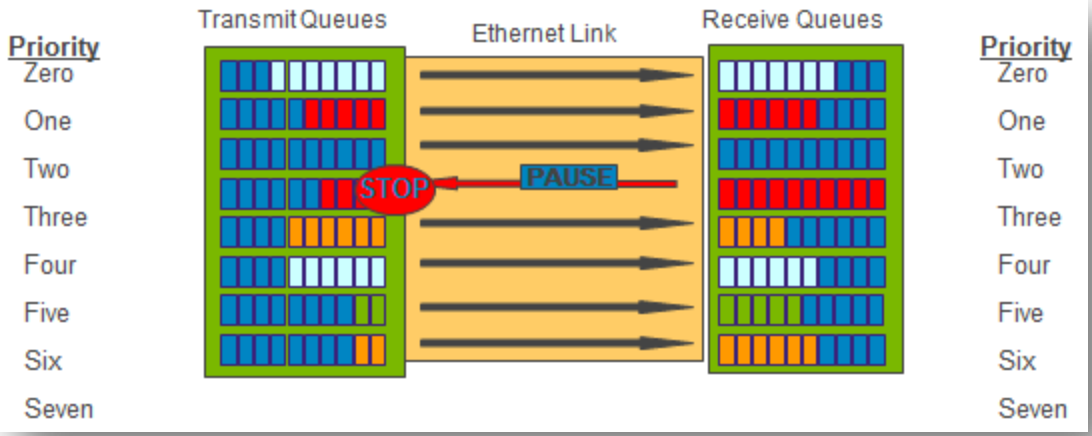

Figure 1. With iSCSI over DCB, frames are paused rather than dropped when the queue is full

- **Enhanced Transmission Selection (ETS) or 802.1Qaz** – Groups multiple classes of service together and then defines a guaranteed minimum bandwidth allocation from the shared network connection
- **Congestion Notification (CN) or IEEE 802.1Qau –** Allows DCB switches to recognize primary bottlenecks and take action to ensure that primary points of congestion do not spread to other parts of the network
- **Datacenter Bridging Capability Exchange (DCBx) –** Helps ensure a consistent configuration across the network, while allowing devices to communicate with each other

Note

Early DCB implementations were typically associated with Fibre Channel over Ethernet (FCoE) rather than iSCSI).

How best to integrate an iSCSI SAN?

A challenge faced in many of today's datacenters is how to integrate an iSCSI SAN with VMware vSphere into an existing network infrastructure. Do you really need to configure four switches: redundant switches for network with additional redundant switches dedicated to iSCSI traffic?

You must decide how to guarantee the integrity of data packets on the storage network. While network traffic may be able to sustain packet losses, such losses are unacceptable for a storage network in a production environment. It has been – and, for many installations, continues to be – a best practice to isolate network traffic from storage traffic by using virtual LANs (VLANs) on one or more network switches. This scenario not only increases cost and complexity but also leads to bandwidth contention that can affect virtual machine (VM) traffic, as well as traffic associated with VMware features like vMotion and Fault Tolerance (FT).

By moving to a converged infrastructure you eliminate the need for multiple, dedicated core switches; now you can use just two redundant switches to carry both network and iSCSI SAN traffic. Moreover, Brocade and Dell switches that provide the appropriate support allow you to implement iSCSI over DCB, creating no-packet-loss capability. As a result, when used with 10GbE network interfaces, an iSCSI storage array is able to perform in much the same way as a Fibre Channel array.

What does iSCSI over DCB provide?

The benefits of iSCSI over DCB can be significant, especially when you introduce lossless connectivity to iSCSI storage solutions such as the EqualLogic PS Series arrays.

It is a best practice with 1GbE networks for storage administrators to separate storage network traffic from the data network, which avoids traffic collisions and, thus, packet loss. Now, however, with Emulex converged network adapters (CNAs) and DCB-supported switches, you can isolate network traffic from storage traffic within a single switch, and then shape bandwidths based on the needs of your workloads. Furthermore, a converged infrastructure solution reduces cable sprawl and lowers your power and cooling costs. Additional benefits are described below.

Leveraging 10GbE

While greenfield datacenters are standardizing on top-of-the rack 10GbE switches, many traditional datacenters are still running 1GbE networks. Although these legacy implementations are viable solutions for network traffic, they are not preferred for iSCSI storage traffic due to the potential for dropped packets, as well as issues with latency.

The bandwidth delivered by 10GbE not only provides a wider data pipe but also gives you the ability to support multiple data pipes. Thus, OCe11102-IM 10GbE adapters can drive significant performance gains by separating I/O and NIC traffic, thus maintaining consistent storage performance even when LAN traffic is varying.

More VMs supported

Emulex conducted performance tests to evaluate VM scalability in test environments that featured the following components:

- OCe11102-IM iSCSI adapter
- 10GbE NIC with software iSCSI

Emulex compared the maximum number of VMs that could run concurrently at a particular I/O rate – a number that was reached in each scenario when I/O throughput dropped below the specified rate. Test results indicated that, for both 4kB and 8kB block sizes, an average of 56 percent more VMs was supported with the iSCSI adapter.

Cost-effectiveness

The arithmetic is simple: it is less expensive to deploy a few 10GbE DCB-enabled ports than a large number of 1GbE non-DCB ports. Furthermore, less overall cable length also translates to cost savings.

Based on the benefits described above, it is easy to see that, with careful testing and planning, the solution described in this document could evolve into an efficiently-performing iSCSI SAN.

iSCSI configurations today

Many of today's datacenters still isolate iSCSI traffic on a separate network with a dedicated switch, while network traffic for management and VMs goes through a second switch. This implementation is so pervasive, it has typically been regarded as a best practice.

Figure 2 shows the hardware needed for a conventional implementation with traffic separation.

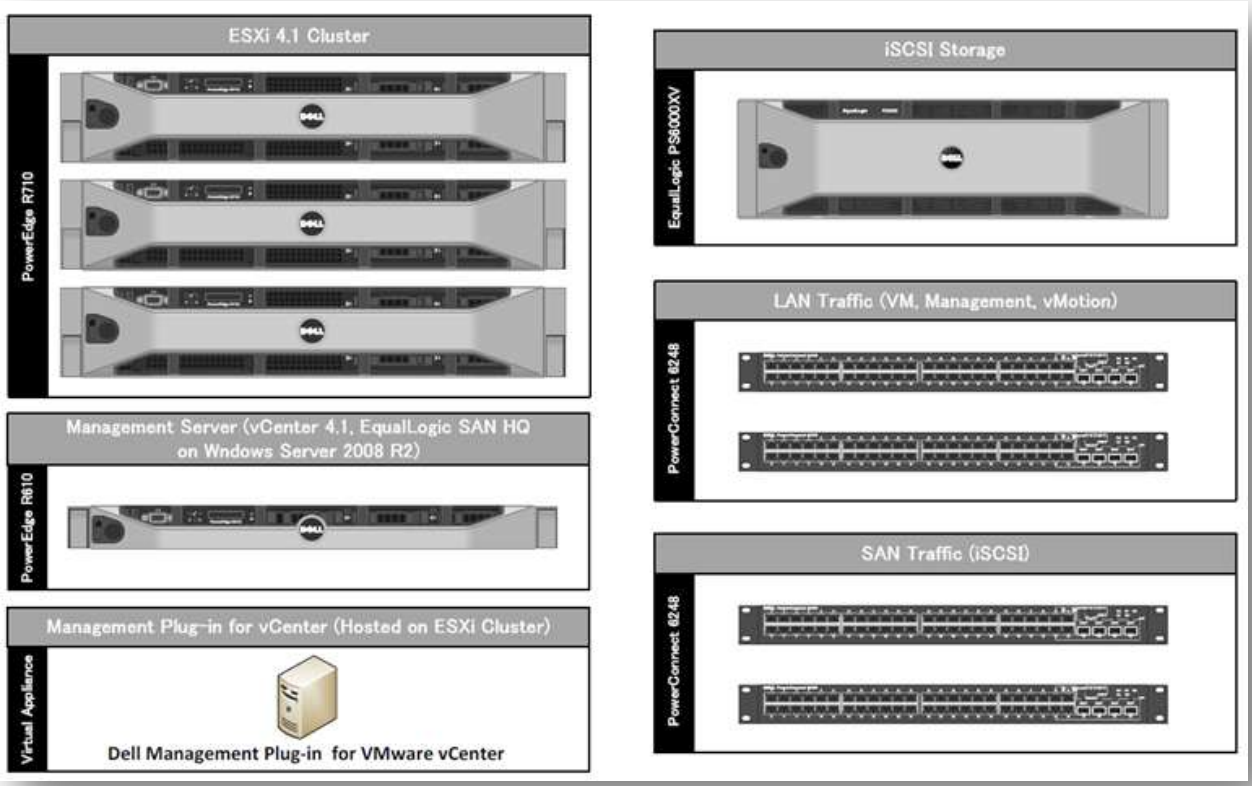

Figure 2. Hardware needed for a network implementation with traffic separation

However, in a converged network infrastructure – for example, based on Dell PowerConnect® switches with Emulex enterprise iSCSI adapters – separate switches are no longer required, as shown in Figure 3.

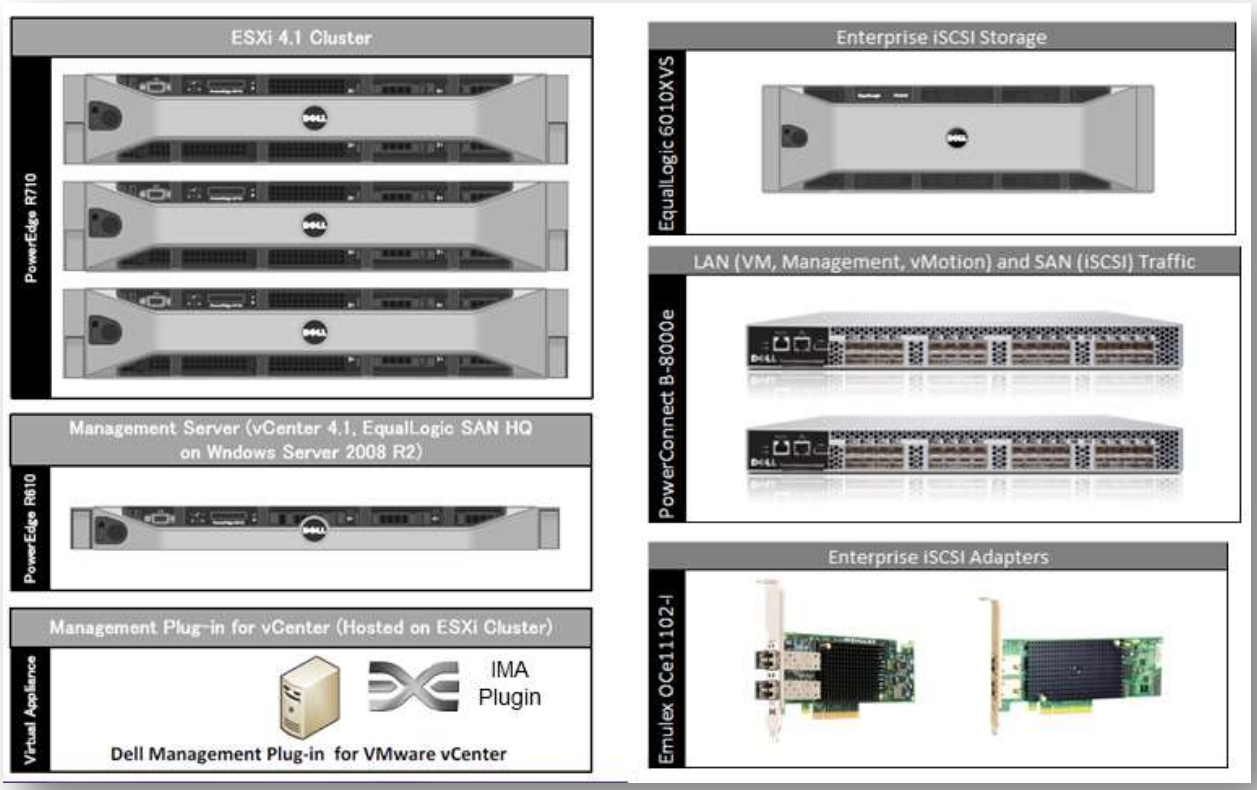
ESXi 4.1 Cluster
PowerEdge R710
Management Server (vCenter 4.1, EqualLogic SAN HQ on Wndows Server 2008 R2)
PowerEdge R810
Management Plug-in for vCenter (Hosted on ESXi Cluster)
Virtual Appliance
IMA Plugin
Dell Management Plug-in for VMware vCenter
Enterprise iSCSI Storage
EqualLogic 6010XVS
LAN (VM, Management, vMotion) and SAN (iSCSI) Traffic
PowerConnect B-8000e
Enterprise iSCSI Adapters
Emulex OCe11102-I

Figure 3. Hardware needed for a converged network implementation

With iSCSI over DCB, there is no separation: iSCSI, management and VM traffic all use the same switch. Note, however, that a second switch is typically provided for redundancy.

Configuring iSCSI over DCB

Important

The iSCSI over DCB proof-of-concept solution described in this section conforms to the appropriate PVSP, which states that the solution is not directly supported by VMware. Thus, any configuration-related issues should be addressed with Dell or Emulex.

In this proof-of-concept, the implementation of iSCSI over DCB takes place in hardware, primarily within the network switch. In a hardware iSCSI implementation (as shown in Figure 4), an Emulex OneConnect adapter can manage all iSCSI, TCP/IP and driver traffic, offloading many associated tasks from the CPU. VMs can connect to the iSCSI storage just as another storage adapter might connect to a SAN. There is no need to create and bind addition vmkernel ports as you would with a software iSCSI storage solution.

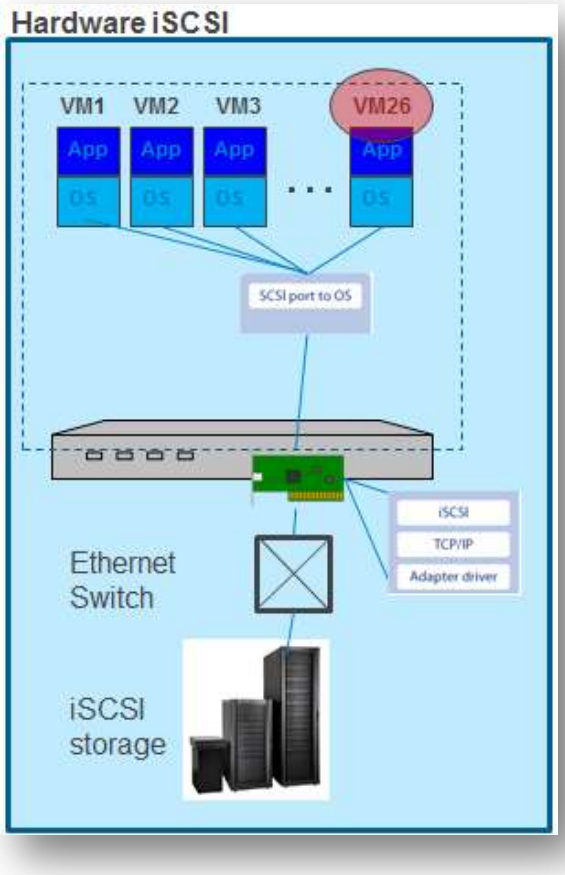

Figure 4. In a hardware iSCSI implementation, an Emulex OneConnect adapter can manage all iSCSI, TCP/IP and driver traffic.

The process for implementing a converged network begins with careful pre-planning. For example, selecting the correct network switch, 10GbE converged network adapter and iSCSI storage array were critical for the proof-of-concept described in this document. In addition, all the hardware used was checked against the VMware Compatibility Guide, which is a good practice for any proof-of-concept involving VMware software.

Note

VMware does not have a specific certification program or compatibility guide for iSCSI over DCB.
For more information on support that may be available from VMware, refer to VMware KB Article 2005240.

The following components were configured for this iSCSI over DCB proof-of-concept solution:

- Dell PowerConnect B-8000e Switch
- Dell EqualLogic PS6010 Storage Array
- Emulex OCe11102-IM iSCSI adapter

Figure 5 shows the proof–of-concept deployment.

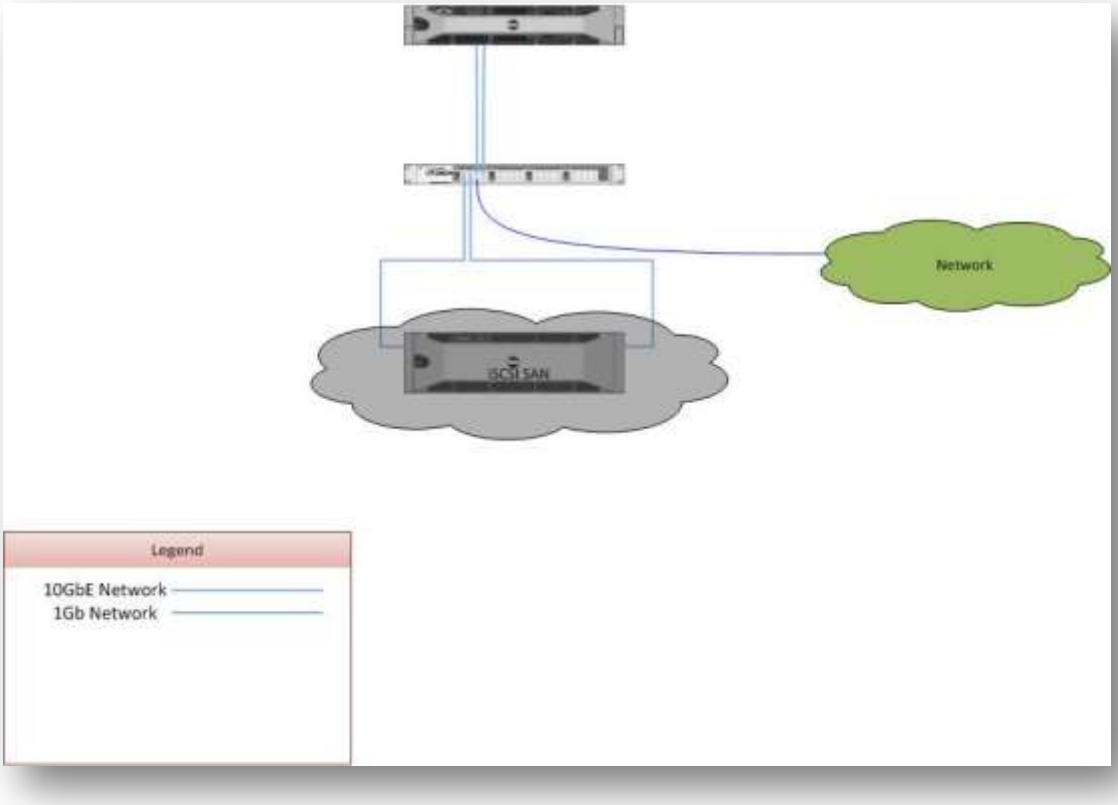

Figure 5. Proof-of-concept for a single-switch iSCSI over DCB deployment with an EqualLogic PS6010 array

The majority of the configuration for iSCSI over DCB takes place within the network switch, except for adding VLAN information. The following sections provide basic information on configuring the network switch and adapter ports for DCB and iSCSI so that array is able to recognize the iSCSI adapter.

Configuring the network switch

Note

Configuring a network switch in your environment may require the assistance of the SAN and network administrator(s).

Both Dell and Brocade offer network switches that support DCB. In this proof-of-concept, Emulex used a Brocade 8000 network switch.

The proof-of-concept outlined in this document was based on a single network switch; as a result, descriptions refer to a single-switch deployment. As a best practice, however, dual-switch deployments are recommended for the datacenter to provide redundancy. Moreover, a Link Aggregation Group (LAG) should be used to interconnect the redundant switches.

Note

As with any other hardware deployment, make sure the latest firmware releases have been certified for use with the PowerConnect B-8000e switch. Contact Dell or Emulex for the latest supported firmware.

The first step when configuring the network switch is to define the types of traffic carried over the CEE network. In the proof-of-concept, traffic was prioritized as follows:

- iSCSI traffic was associated with priority 3, which was grouped into a priority group (PG) named **PGID1**
- IP traffic was associated with priorities 0 – 2 and 4 – 7, which were grouped into a PG named **PGID2**

Note

Emulex Ethernet adapters support up to two PGs – PGID1 and PGID2 in this example.

For a sample switch configuration output, refer to Appendix B – Network switch command line configuration.

Configuring the storage array

Since EqualLogic PS Series iSCSI arrays are ideal for a vSphere 5 implementation, a PS6010 array was used in the proof-of-concept.

By default, DCB is enabled on all PS Series arrays running firmware version 5.1.0 or later. To support an iSCSI over DCB implementation, you only need to configure the VLAN ID for DCB.

Figure 6 provides a view of EqualLogic PS Group Manager, showing the **Group Configuration** pane's **Advanced** tab, which allows you to enable DCB.

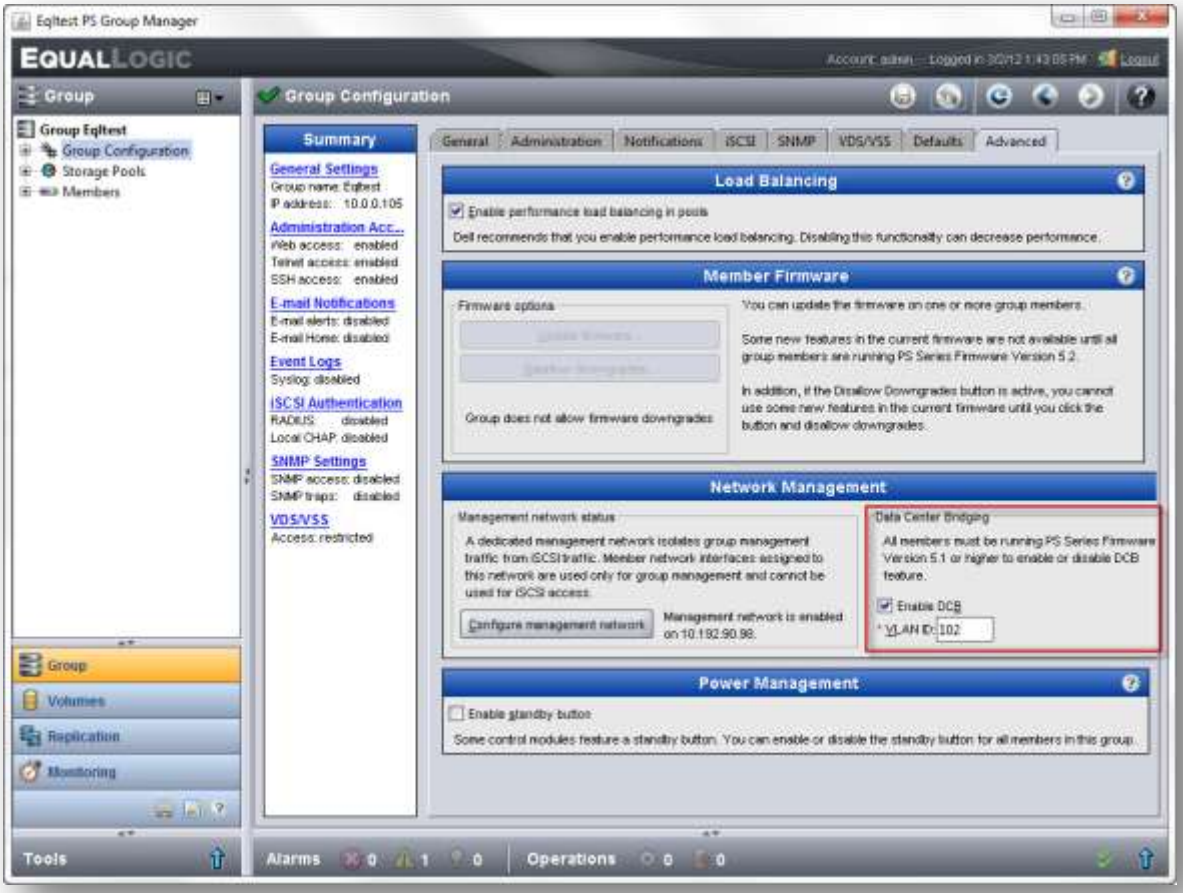

Figure 6. Configuring DCB via the Group Configuration pane of PS Group Manager

Once the PS6010 array has been integrated into the SAN, it determines if should operate in standard or DCB Ethernet mode based on the particular switch port settings that have been configured.

Emulex OneConnect® OCe11100 adapters

Emulex OneConnect series adapters have been providing support for PFC in a vSphere environment since the introduction of the Emulex LP21000 CNA. Now, the OCe11100 family of adapters provides support for NIC, FCoE and iSCSI traffic, making these devices true CNAs.

vSphere 5 provides inbox drivers for network functionality but no iSCSI driver, which must be downloaded from the VMware website and installed manually or added via VMware Update Manager.

Before configuring the OCe11100 adapter, verify that the latest firmware has been uploaded to the adapter. You can use Emulex OneCommand® Manager or VMware vCenter Server plug-in for OneCommand Manager to check the installed version. If necessary, manually download the latest version.

In the proof-of-concept, the following steps were used to configure the iSCSI adapter via the Emulex iSCSISelect Utility tool (included in firmware version v4.0.360.3 or later):

1. Install the adapter on the vSphere host and boot the server.

2. Via the BIOS screen for the iSCSI adapter, select the keyboard combination <**CTRL**> **S** to invoke the iSCSISelect Utility tool, as shown in Figure 8:

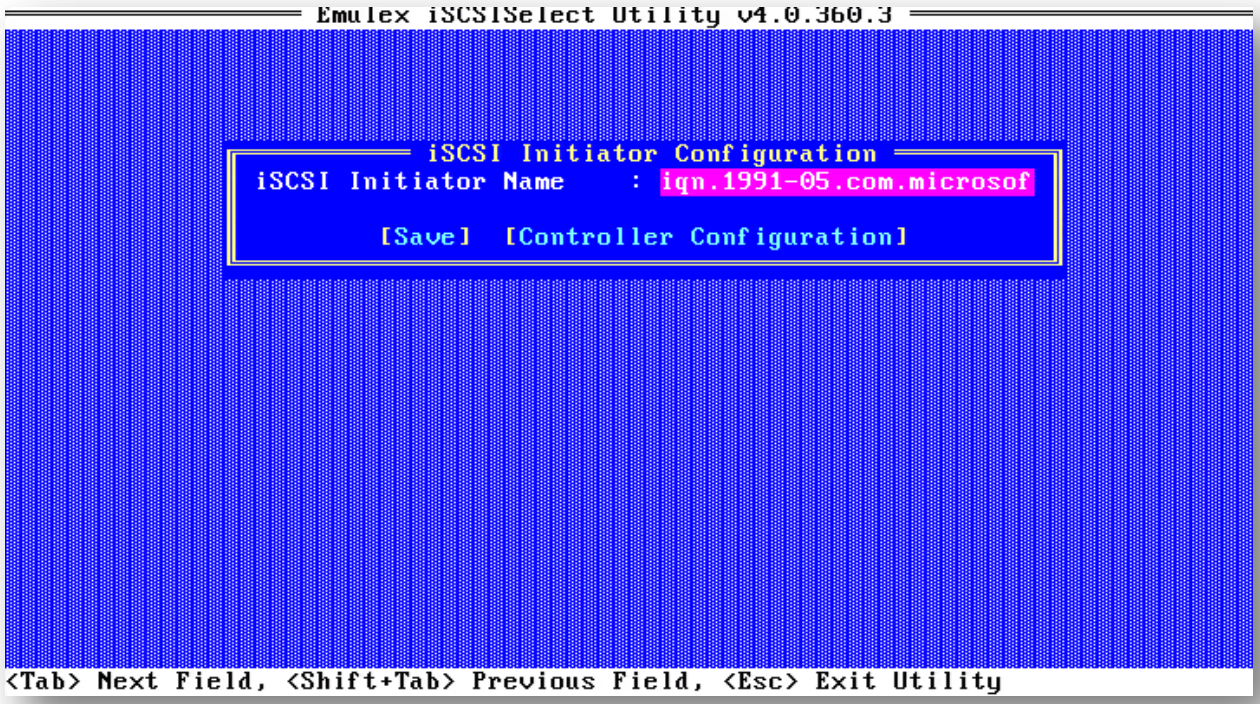
Emulex iSCSISelect Utility v4.0.360.3
iSCSI Initiator Configuration
iSCSI Initiator Name : iqn.1991-05.com.microsof
[Save] [Controller Configuration]
<Tab> Next Field, <Shift+Tab> Previous Field, <Esc> Exit Utility

Figure 8. Invoking iSCSISelect Utility to configure the iSCSI adapter

3. Select the controller[1] port to be configured. For the purposes of this proof-of-concept, Emulex configured a single port (as shown in Figure 9); however, in most cases, you would configure a second port just like the first to provide redundancy/load balancing capabilities.

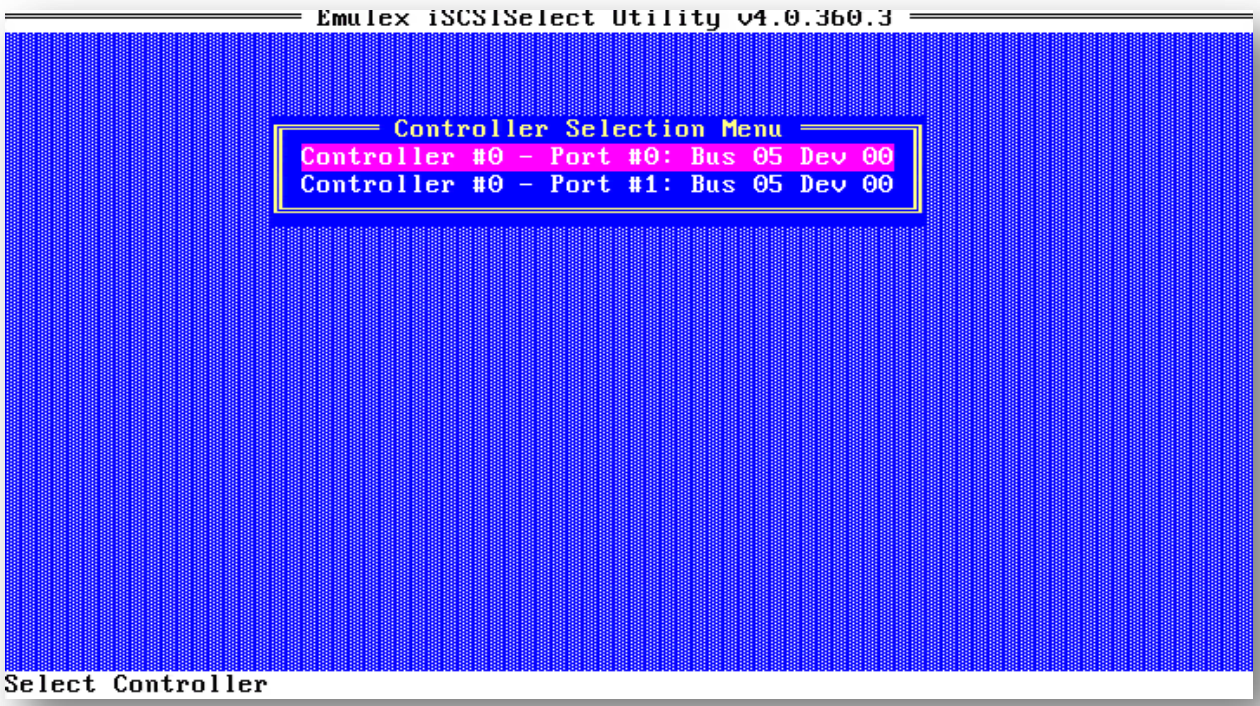
Emulex iSCSISelect Utility v4.0.360.3
Controller Selection Menu
Controller #0 - Port #0: Bus 05 Dev 00
Controller #0 - Port #1: Bus 05 Dev 00
Select Controller

Figure 9. Selecting the controller port

4. Start configuring the selected controller port.

   Use **Network Configuration** to select **Port 0**.

   In most cases you should disable DHCP and use a static IP address. If the IP address lease were to renew, it may be difficult to log into the target.

   Select **Configure VLAN ID/Priority** and enable VLAN support if you plan to deploy or join a network with VLANs; otherwise, disable this option.

   Select **Save** to exit.
5. Enter the **Static IP address**, **SubnetMask** and **Default Gateway**, then select **Save** to exit.
6. Select the option to test network connectivity via the Ping target utility.

[1] Here, “controller” refers to the iSCSI adapter.

7. Select the **iSCSI Target Configuration** option, then **[Add New iSCSI Target]**, as shown in Figure 10.

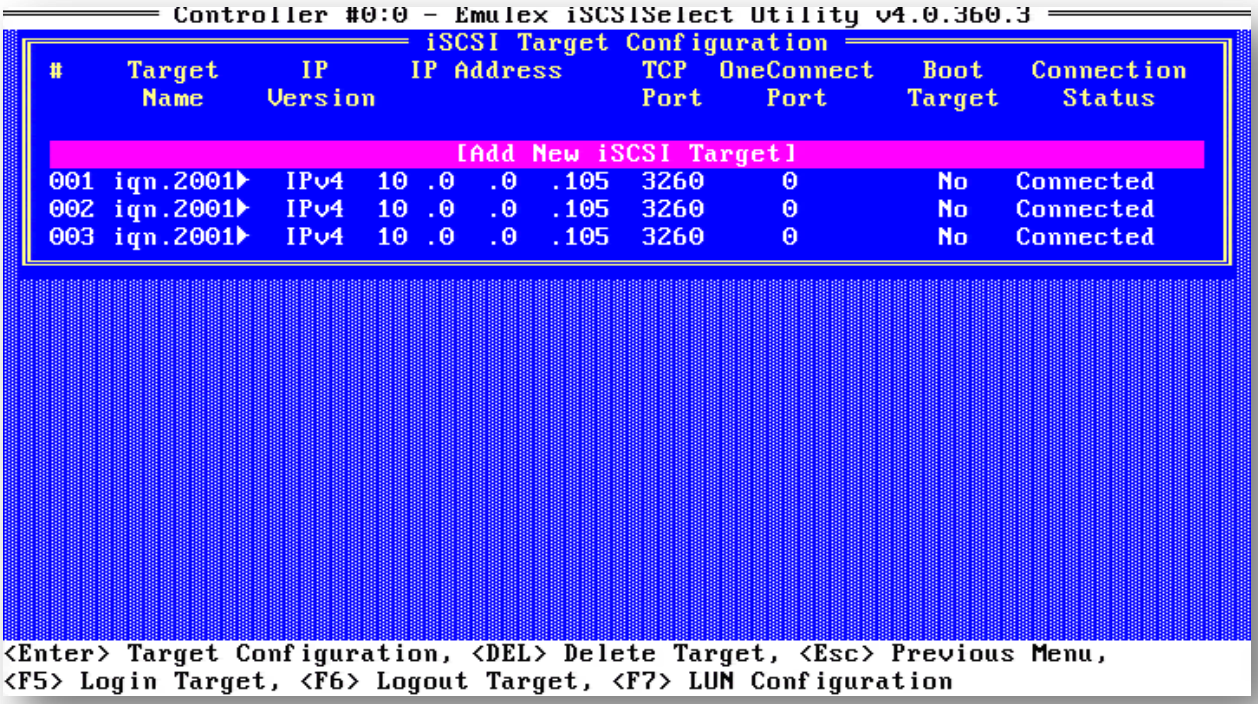
Controller #0:0 - Emulex iSCSISelect Utility v4.0.360.3
iSCSI Target Configuration
Target Name
IP Version
IP Address
TCP Port
OneConnect Port
Boot Target
Connection Status
[Add New iSCSI Target]
001 iqn.2001 IPv4 10 .0 .0 .105 3260 0 No Connected
002 iqn.2001 IPv4 10 .0 .0 .105 3260 0 No Connected
003 iqn.2001 IPv4 10 .0 .0 .105 3260 0 No Connected
<Enter> Target Configuration, <DEL> Delete Target, <Esc> Previous Menu,
<F5> Login Target, <F6> Logout Target, <F7> LUN Configuration

Figure 10. Adding the new iSCSI target

8. Enter information such as **Target Name**, **IP address** and **TCP Port** number.

Note

Consult the SAN administrator, if appropriate.

Once you have finished configuring the iSCSI adapter, network switch and storage array, the adapter ports should be able to log in to the targets presented. You can then install vSphere 5 on local storage and begin deploying VMs on the shared iSCSI storage.

Validation and troubleshooting

In order to validate the configuration of the proof-of-concept solution, Emulex used Iometer, an easy-to-use tool that is able to generate a range of workloads. Other techniques are available but are beyond the scope of this document.

This section outlines how to validate a DCB over iSCSI solution and provides some guidelines that may be useful when troubleshooting connectivity issues.

Validating the iSCSI adapter

To verify that the iSCSI adapter has been configured correctly, you should install either OneCommand Manager or the OneCommand Manager for VMware vCenter plug-in (as in Figure 11).

Use the **Emulex OneCommand** tab of vCenter to verify that the iSCSI adapter port is connected to a target. Select the **Initiator** view and then the **iSCSI Target Discovery** button; status should show as **Connected**.

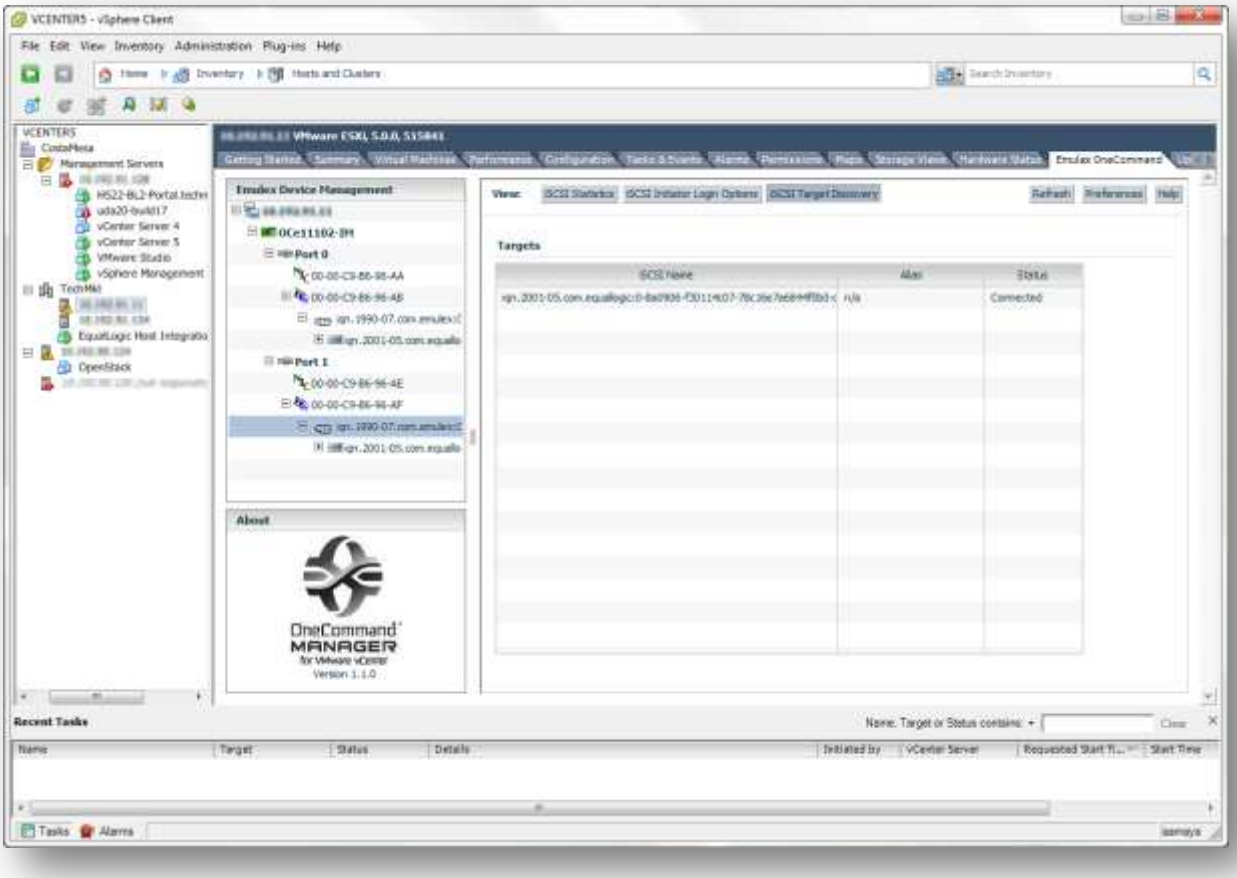

Figure 11. Verifying that Port 1 is connected to a target

Obtaining link status

The OneCommand Manager for VMware vCenter plug-in can also provide event status for links. Thus, if you are unable to log in and start a session with the array, connect to vCenter and select the **Tasks & Events** tab, followed by the **Events** button. In the example shown in Figure 12, Emulex artificially created errors on both ports to demonstrate what happens when a link goes down.

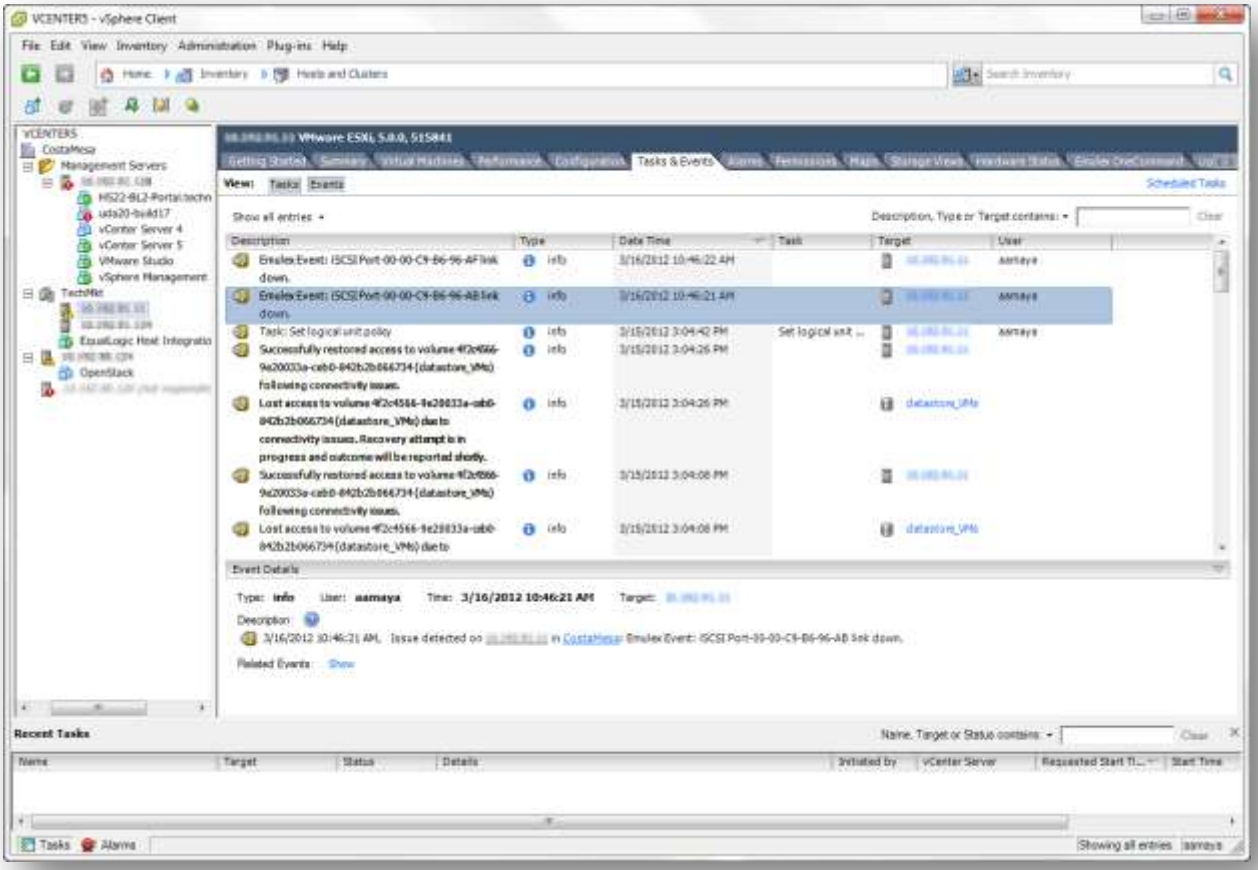

Figure 12. View of artificially-created link-down events

Testing the network switch

The following commands may be useful when troubleshooting switch connectivity (Port 0 in these examples):

- Clearing LLDP neighbor information:

  #clear lldp neighbors tengigabitethernet 0/0

- Clearing LLDP statistics:

  #clear lldp statistics tengigabitethernet 0/0

- Displaying LLDP neighbors:

  #show lldp

- Displaying LLDP interface information:

  #show lldp interface tengigabitethernet 0/0

- Displaying LLDP neighbor-related information

  #show lldp neighbors interface tengigabitethernet 0/0

Conclusion

After a Dell EqualLogic PS Series array has been configured for iSCSI over DCB in conjunction with an Emulex OneConnect OCe11102-IM iSCSI adapter, vSphere 5 recognizes this set-up in the same way as any other hardware iSCSI adapter connected to iSCSI storage. However, although Emulex had to enable VLANs, set the bandwidth and enable priority flow control, most configuration for DCB over iSCSI occurs at the switch level and is transparent to vSphere 5. Furthermore, the Dell PowerEdge R710 server used in the proof-of-concept did not require any special configuration for DCB support – indeed, any Dell server listed in the VMware Compatibility Guide could have been used.

As with any hardware deployment, Emulex stresses the importance of thoroughly researching an iSCSI over DCB solution, and assessing not only the benefits but also potential risks.

Following a successful solution deployment, the next stage of the lifecycle is management. Both Dell and Emulex offer plug-ins for VMware vCenter Server 5, resulting in single-glass pane of management for storage and the iSCSI adapter.

Appendix A- Bill of materials

Table A-1. Components used in the proof-of-concept

Product	Description
Server	Dell PowerEdge R710
Memory	24GB RAM
Network	On-board 1Gb
iSCSI hardware	Emulex OCe11102-IM Dual Port iSCSI Adapter
Disks	SAS, 146GB
RAID	Dell Perc 6/i with 256BM battery-backed cache
Software	VMware ESXi 5.0 Enterprise Edition, vSphere Client and Emulex OneCommand Manager plug-in for vCenter Server

Appendix B – Network switch command line configuration

In the proof-of-concept, Emulex configured eight ports on the network switch (ports 0 – 7) in order to validate the configuration and provision a LUN. To support these capabilities, it was critical for all the ports to be configured identically .

Because a hardware iSCSI adapter was being used, maximum transmission unit (MTU) settings for each port were changed.

Configuration activities were as follows:

- Set the MTU size for ports 0 – 7 to 9208
- Configure the switch for Rapid Spanning-Tree
- Enable edge ports
- Set ports 0 – 7 to layer 2 converged mode
- Apply the default Converged Enhanced Ethernet (CEE) map
- Define the iSCSI priority class

The following output reflects the switch configuration:

Note

Ports 8 – 24 are not shown since they were not used in the testing.

```
Brocade8K#show running-config
!
protocol spanning-tree rstp
!
cee-map default
      priority-group-table 0 weight 50
      priority-group-table 1 weight 50
      priority-table 0 0 0 0 1 0 0 0
interface TenGigabitEthernet 0/0
      mtu 9208
      switchport
      switchport mode converged
```

```
      switchport converged allowed vlan all
      no shutdown
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default
!
interface TenGigabitEthernet 0/1
      mtu 9208
      switchport
      switchport mode converged
      switchport converged allowed vlan all
      no shutdown
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default
!
interface TenGigabitEthernet 0/2
      mtu 9208
      switchport
      switchport mode converged
      switchport converged allowed vlan all
      no shutdown
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default
```

```
!
interface TenGigabitEthernet 0/3
      mtu 9208
      switchport
      switchport mode converged
      switchport converged allowed vlan all
      no shutdown
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default
!
interface TenGigabitEthernet 0/5
      mtu 9208
      switchport
      switchport mode converged
      switchport converged allowed vlan all
      no shutdown
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default
!
interface TenGigabitEthernet 0/6
      mtu 9208
      switchport
      switchport mode converged
      switchport converged allowed vlan all
      no shutdown
```

```
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default
!
interface TenGigabitEthernet 0/7
      mtu 9208
      switchport
      switchport mode converged
      switchport converged allowed vlan all
      no shutdown
      lldp iscsi-priority-bits list 4
      spanning-tree edgeport
      cee default

(Ports 8-24 not configured)
       protocol lldp
       system-description Brocade 8000
       advertise dcbx-fcoe-app-tlv
       advertise dcbx-iscsi-app-tlv
       advertise dcbx-fcoe-logical-link-tlv
```

```
!
line console 0
       login
line vty 0 31
       login
!
end
Brocade8K#
```

Appendix C- Emulex OneCommand OCe11102-IM adapter configuration

Table C-1. Emulex OCe11102-IM iSCSI adapter configuration

Component	Level or setting
Firmware	4.0.360.3
iSCSI driver	Be2iscsi 4.0.317.0
VLAN	1
IP Address	10.0.0.105
IP Address	10.0.0.205

For more information

Dell white paper: “EqualLogic PS Series Reference Architecture for PowerConnect B-Series 8000 and 8000e”	http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cts=1330644034144&ved=0CC0QFjAB&url=http%3A%2F%2Fen.community.dell.com%2Fdell-groups%2Fdtcmedia%2Fm%2Fequallogic%2F19919713%2Fdownload.aspx&ei=CANQT-6QMavRiAK4pvW0Bg&usg=AFQjCNH6PMeszL46PpS-i7iUHKx6I1R2vw&sig2=_8YV1bY2MWqwOizv3rDgiQ
Dell Storage wiki: “Creating a DCB Compliant EqualLogic iSCSI SAN with Mixed Traffic”	http://en.community.dell.com/techcenter/storage/w/wiki/creating-a-dcb-compliant-equallogic-iscsi-san-with-mixed-traffic.aspx
VMware Knowledge Base article: “Dell and Emulex iSCSI over DCB solution for VMware vSphere (Partner Verified and Supported)”	http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2005240
Emulex OneConnect iSCSI adapters	http://www.emulex.com/products/10gbe-iscsi-adapters.html
Emulex Implementer’s Lab	http://www.emulex.com/the-implementers-lab.html

To help us improve our documents, please provide feedback at implementerslab@emulex.com.

Some of these products may not be available in the U.S. Please contact your supplier for more information.

© Copyright 2012 Emulex Corporation. The information contained herein is subject to change without notice. The only warranties for Emulex products and services are set forth in the express warranty statements accompanying such products and services. Emulex shall not be liable for technical or editorial errors or omissions contained herein.
OneConnect and OneCommand are registered trademarks of Emulex Corporation. Dell is a registered trademark in the U.S. and other countries.
VMware is a registered trademark of VMware Corporation.

EMULEX

www.emulex.com

World Headquarters 3333 Susan Street, Costa Mesa, California 92626 +1 714 662 5600
Bangalore, India +91 80 40156789 | **Beijing, China** +86 10 68499547
Dublin, Ireland+35 3 (0)1 652 1700 | **Munich, Germany** +49 (0) 89 97007 177
Paris, France +33 (0) 158 580 022 | **Tokyo, Japan** +81 3 5322 1348
Wokingham, United Kingdom +44 (0) 118 977 2929